JOURNAL OF PALLIATIVE MEDICINE
Volume 12, Number 1, 2009

DOI: 10.1089/jpm.2008.9690

Study Design, Precision, and Validity in Observational Studies

Melissa D.A. Carlson, Ph.D.[1,2] and R. Sean Morrison, M.D.[1–3]

Introduction

The use of observational research methods in the field of palliative care is vital to building the evidence base, identifying best practices, and understanding disparities in access to and delivery of palliative care services. As discussed in the introduction to this series, research in palliative care encompasses numerous areas in which the gold standard research design, the randomized controlled trial (RCT), is not appropriate, adequate, or even possible.[1,2] The difficulties in conducting RCTs in palliative care include patient and family recruitment, gate-keeping by physicians, crossover contamination, high attrition rates, small sample sizes, and limited survival times. Furthermore, a number of important issues including variation in access to palliative care and disparities in the use and provision of palliative care simply cannot be answered without observational research methods. As research in palliative care broadens to encompass study designs other than the RCT, the collective understanding of the use, strengths, and limitations of observational research methods is critical. The goals of this first paper are to introduce the major types of observational study designs, discuss the issues of precision and validity, and provide practical insights into how to critically evaluate this literature in our field.

Observational Studies

Observational studies draw inferences about the effect of an "exposure" or intervention on subjects, where the assignment of subjects to groups is observed rather than manipulated (e.g., through randomization) by the investigator. Observational research involves the direct observation of individuals in their natural setting. As such, who does or does not receive an intervention is determined by individual preferences, practice patterns, or policy decisions.[3] It is therefore important for readers of observational research to consider if alternative explanations for study results exist. This issue (known as "confounding") is a primary challenge of observational research and will be discussed in detail in the next paper in this series.

Data for observational research is either collected by the investigator for the purpose of the study (primary data) or has already been collected for another purpose but is used by the investigator to examine a novel research question (secondary data). The primary trade-offs between using primary and secondary data relate to time, resources, and control of the collection and measurement of study variables (Table 1).

A common source of secondary data used for observational research is administrative data. For example, data from Medicare claims allow researchers to study the health care utilization of large groups of individuals. Studies using Medicare claims data are observational because the investigator is observing the subjects' health care utilization without any contact or involvement with the subjects. It is secondary data because the data were collected by the Centers for Medicare and Medicaid Services for purposes other than the investigators' study. Other sources of data typically used for observational research include hospital administrative databases, data obtained from medical chart review, or data obtained from previously conducted research studies.

Types of Observational Study Designs

There are three main types of observational study designs that are distinguished by the objective of the research study, how subjects are sampled, and the timeline of data collection. In evaluating and critically appraising observational studies, it is important for readers to consider if the study design was appropriate for the research question and if the methodology used was consistent with the study design. A comparison of experimental and observational study designs is shown in Table 2.

Cross-sectional studies

A cross-sectional study is an observational study in which exposure and outcome are determined simultaneously for each subject. It is often described as taking a "snapshot" of a group of individuals. Cross-sectional studies are most ap-

[1]National Palliative Care Research Center, New York, New York.
[2]The Hertzberg Palliative Care Institute of the Brookdale Department of Geriatrics, Mount Sinai School of Medicine, New York, New York.
[3]James J. Peters Veterans Affairs Medical Center, Bronx, New York.

Table 1. Advantages and Disadvantages of Using Primary and Secondary Data

Primary Data	Secondary Data
Advantages	*Advantages*
Investigator controls all aspects of the study including design, sampling techniques, data collection, and follow-up methods	Relatively fast and inexpensive
All variables of interest can be measured	Sample sizes tend to be large
	Sample may cover a large geographic area and thus provide ability to assess national trends
	Unobtrusive to subjects
Disadvantages	*Disadvantages*
Time consuming	Data may not include all variables of interest
Expensive	Maybe difficult to understand how and why data elements were collected

propriate for screening hypotheses because they require a relatively shorter time commitment and fewer resources to conduct.[5]

Cross-sectional studies have been widely used in palliative care research. The cross-sectional study design has been used to understand the prevalence of various conditions, treatments, services or other outcomes and the factors associated with such outcomes. For example, we have used a cross-sectional study[7] to identify the specific services provided to patients who enrolled with hospice and the extent to which services varied across hospices. The cross-sectional study design was an efficient way to evaluate a large sample of patients receiving hospice, to understand the prevalence of specific services, and to generate hypotheses regarding why service delivery might vary across hospices. Similarly, we used a cross-sectional design to estimate the association between hospice ownership and the provision of specific types of hospice services.[8] Other examples of the use of cross-sectional designs in palliative care research include a study of the association between caregiver characteristics (e.g., sociodemographics, the existence of social networks) and caregiver burden among caregivers of terminally ill patients[9] and the association between physician characteristics (e.g., age, gender, specialty, board certification, knowledge about hospice) and referral of patients to hospice.[10]

Issues that a reader should consider in evaluating a cross-sectional study are threefold. First, the primary limitation of the cross-sectional study design is that because the exposure and outcome are simultaneously assessed, there is generally no evidence of a temporal relationship between exposure and outcome. That is, although the investigator may determine that there is an association between an exposure and an outcome, there is generally no evidence that the exposure caused the outcome. Of course, if the exposure is a characteristic such as gender or race and the outcome developed over time, the temporal nature of the exposure-outcome as-

Table 2. Comparison of Experimental and Observational Study Designs[4–6]

	Experimental	*Observational*		
Study design	Randomized Control Trial	Cross-sectional	Cohort	Case-control
Study population	Highly selected population; highly controlled environment	Diverse population observed in a range of settings	Diverse population observed in a range of settings	Diverse population observed in a range of settings
Directionality	Exposure is assigned before outcome is ascertained	Exposure and outcome ascertained simultaneously	Exposure is ascertained before outcome is ascertained	Outcome is ascertained before exposure is ascertained
Primary Use	Demonstrating efficacy of an intervention	Screening hypotheses; prevalence studies	Assessing association between multiple exposures and outcomes over time	Assessing associations between exposures and rare outcomes
Analysis	Straight-forward	Sophisticated multivariate techniques may be required to account for confounding	Sophisticated multivariate techniques may be required to account for confounding	Sophisticated multivariate techniques may be required to account for confounding
Internal validity	High	Low	Low	Low
External validity	Low-Moderate	High	High	High

sociation is more plausible; however, for studies in which the exposure is not an inherent trait but one that developed over time, causality is often unclear. Second, a cross-sectional study evaluates prevalent rather than incident outcomes and thus excludes people who develop the outcome but die before the study. The measured association in a cross-sectional study is between exposure and having the outcome as opposed to exposure and developing the outcome. As such, there is a bias toward including in the study individuals with more favorable survivorship.[5] For example, early cross-sectional studies that observed beneficial effects of postmenopausal estrogen use on cardiovascular disease in older women failed to account for the increase in cardiovascular-related mortality that, as a result of randomized studies that followed women over time, we now know occurs within the first several years of hormone replacement therapy.[11] Third, the reader needs to assess if alternative explanations for study results have been appropriately ruled out.

Issues to Consider: Cross-Sectional Study

1) If causality is asserted, is it appropriate?
2) Is the distinction between prevalent and incident cases a factor?
3) Have alternative explanations been ruled out?

Cohort studies

The identifying feature of a cohort study design is that the subjects are followed over time. Cohort studies begin with individuals who are exposed and not exposed to a factor and then evaluate the subsequent development of an outcome. Cohort studies may be concurrent or retrospective, the distinction being when, relative to the current time, the subjects are identified (Fig. 1). Cohort studies are an appropriate study design when: (1) there is good evidence to suggest an association between an exposure and an outcome (perhaps through prior cross-sectional studies); (2) the interval between exposure and development of the outcome is relatively short to minimize loss to follow-up; and (3) the outcome is not too rare (so that the size of the cohort is reasonable). The advantages of the cohort study design are that because the investigator identifies new or "incident" cases of the outcome, one can look at disease progression, staging, and natural history. Cohort designs can yield incidence rates as well as relative risks, and cohort studies may be able to assess causality due to the temporal nature of the study design.

In palliative care, cohort studies have been most useful in evaluating the effect over time of palliative care interventions. For example, a recent retrospective cohort study[12] evaluated the effect of palliative care consultation on family satisfaction with care. The authors identified a group of family members of patients who had died in the hospital and had received a palliative care consultation prior to death (the "exposed" group) and who had not received a palliative care consultation prior to death (the "not exposed" group). The investigators then contacted the families and administered a telephone survey to ascertain the family's satisfaction with care. The cohort study design enabled the researchers to conclude that hospital palliative care consultation was associated with improved family outcomes. Similarly, in a concurrent (or often called "prospective") cohort study, investigators studied the effect of pain and opioid analgesia on the development of delirium.[13] The study enrolled all patients presenting to the hospital with hip fracture and without delirium and followed them through their hospitalization collecting data on pain, delirium risk factors, and analgesic prescribing. The results of this study demonstrated that untreated pain was a significant risk factor for the development of delirium and that opioid analgesics decreased the risk of developing this condition.

Cohort study designs are increasingly used in palliative care research in what are known as "quasi-experimental" studies. These cohort studies combine elements of observational and experimental research methods. Quasi-experimental designs are similar to experimental designs in that there is a specific investigator-defined intervention for the "exposed" group in the study, but individuals are not randomized to receive the intervention. Individuals are simply observed as having or not having the intervention (or exposure) and outcomes are subsequently assessed. For example, to study the effect of bereavement counseling on caregiver outcomes, an investigator could design a specific bereavement counseling intervention and offer it to each family that receives a palliative care consult at a specific hospital. Some families will choose the intervention and some will not. The investigator then compares outcomes for the families that received and did not receive the bereavement counseling intervention. The aforementioned study is considered quasi-experimental because it involves a specific intervention designed and implemented as part of the study (experimental) but subjects are not randomized to the intervention; rather their receipt or nonreceipt of the intervention is ob-

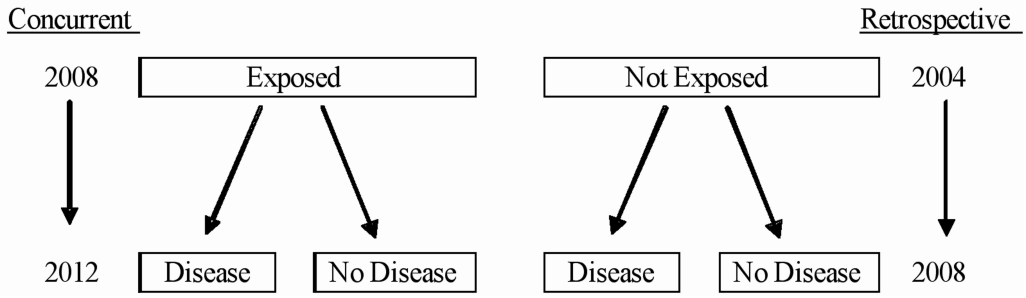

FIG. 1. Cohort study design.

served by the investigator. Quasi-experimental study designs are increasingly used in palliative care research to evaluate the effectiveness of clinical or educational interventions.

There are a number of important issues to consider in evaluating an observational cohort study in palliative care. One issue is loss to follow-up, particularly differential loss to follow-up. Loss to follow-up occurs when, during the study period, individuals drop out of the study. Differential loss to follow-up is when the drop-out rate differs in the exposed and not exposed groups. The concern is that differential loss to follow-up introduces bias into the study. Readers should look for a statement regarding loss to follow-up and whether or not it differed between the study groups. Second, as in cross-sectional studies, the existence of alternative explanations for study results due to confounding must be carefully considered. Third, readers need to assess if there is potential bias in outcome assessment. Given the team-oriented approach to palliative care, it is often difficult to blind (i.e., keep the exposure status of the study participant unknown) the investigators who are assessing the study outcome. Knowing the exposure status of the study participant may influence or bias the assessment of the outcome.

Issues to Consider: Cohort Study

1) Was there significant loss to follow-up?
2) Was loss to follow-up differential?
3) Was there potential bias in outcome assessment?
4) Have alternative explanations been ruled out?

Case-control studies

Case-control studies begin with individuals who have the outcome ("cases") and compare them to individuals who do not have the outcome ("controls") according to past history of exposure to a factor (Fig. 2). Case-control studies are appropriate when: (1) the outcome is rare and (2) there is reliable evidence of past exposure. One issue to consider in ascertaining past exposure is recall bias. Past exposure is generally ascertained by interviewing subjects or analyzing historical records or charts. If cases and controls differentially recall past exposures or there is more or less thorough documentation on cases compared to controls, study results may be biased. For case-control studies, the general concern is that cases will be more likely than controls to recall past exposures because they have already considered the potential causes of their disease. Similarly, interviewer bias occurs when study investigators interview cases more thoroughly regarding past exposures than controls because they know the subject is a case. Readers of case-control studies should consider the potential extent of recall or interviewer bias and whether study investigators attempted to mitigate these issues.

Sample selection in a case-control study is complex. Cases may be selected from a variety of sources including hospital patients, patients in a physician's practice, clinic patients, and cancer registries. It is desirable to select cases from multiple institutions (e.g., multiple hospitals in the community instead of one hospital) to obtain more generalizable results. Criteria for case eligibility should be carefully specified in the Methods section. It is preferable to use incident ("newly diagnosed") cases so that risk factors identified are not related to survival with the outcome as opposed to development of the outcome.

The most important issue to consider in critically evaluating a case-control study is the process by which controls were selected and the resulting comparability of cases and controls. The selection of controls is the most complex and controversial aspect of conducting a case-control study. Controls should be similar to cases in all respects other than having the disease or should be similar to the general population from which the cases arose. Common sources of controls include the spouse, friend, or neighbor of the case, an individual hospitalized at the same time as the case but for a different reason, or an individual chosen randomly from the general population. Controls chosen from the general population used to be often ascertained from random digit dialing. However, the increased use of answering machines, "Do Not Call" lists, and cell phones has rendered random digit dialing a less effective option as individuals reached by land-line phone may no longer be representative of the general population. Some case-control studies use matching, which refers to selecting controls so that they are similar to cases in specific characteristics (e.g., race, age, gender, socioeconomic status). Cases that are unable to be matched are often excluded from the analyses. Readers should determine the proportion of cases that were excluded from the analyses because a high proportion could limit the generalizability of the study.

To date, case-control studies have not been widely used in palliative care research. This is not surprising as the primary benefit of the case-control design is the ability to

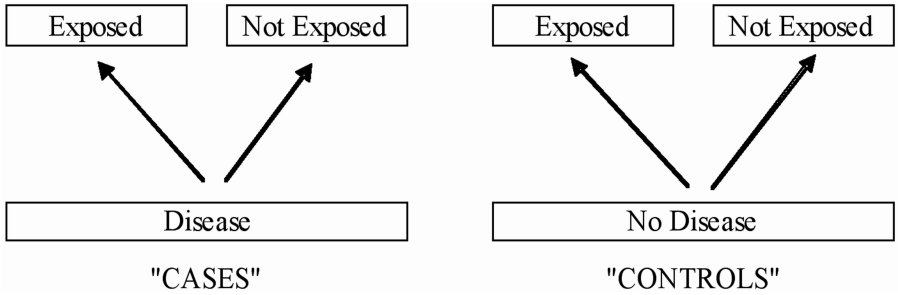

FIG. 2. Case-control study design.

study rare outcomes (e.g., rare diseases) and to look back in time for exposures that may be correlated with the outcome. At this point, palliative care research is generally not focused on studying rare outcomes and thus the benefit of the case-control design is more limited. However, one area of geriatrics research that has used the case-control study design is the understanding of factors associated with hospital falls. An initial understanding of risk factors for falling in the hospital came from a case-control study conducted in a large urban academic hospital.[14] The investigators identified 98 patients who fell while they were inpatients (cases) and compared them to 318 inpatients who did not fall (controls). They then interviewed each patient in the study to assess potential patient-related, medication-related, and care-related risk factors. The case-control study design was ideal in that it allowed the investigators to study a fairly rare outcome and yet obtain a relatively large sample size.

Issues to Consider: Case-Control Study

1) Do controls come from the same hypothetical population as cases?
2) Is recall or interviewer bias an issue?
3) Have alternative explanations been ruled out?

Evaluating the Precision of an Observational Study

In addition to the challenges that readers must consider arising from the specific design of an observational study, there are two additional challenges that apply to observational research of any design. The first challenge is precision. Precision refers to lack of random error or random variation in a study's estimates.[5] In observational studies, random variation arises from the subjects in the study, the way in which subjects are sampled, and the way in which variables are measured. Subjects in a study are always considered a sample of possible individuals who could have been included in the study but were not and thus the sample selection introduces random variation. The measurement of key variables also introduces random variation. Because most observational studies must include potential confounding variables, random variation due to the measurement of these variables will likely exist (this is compared with an RCT in which the randomization may eliminate the need to include potential confounding variables).

As a reader of an observational study, one can get a sense of the precision by considering both the sample size and the efficiency of the study. In general, a larger study and one with more balanced groups (i.e., exposed, not exposed, with outcome, without outcome) will produce more precise estimates.[5] For example, a study with a large sample size ($n = 1000$) but only a small number of subjects who are exposed ($n = 50$) compared with not exposed ($n = 950$) will yield less precise estimates than a smaller sample size where roughly half of participants are exposed and not exposed. Similarly, the proportion of subjects with the outcome and the distribution of subjects across key covariates will impact the efficiency and thus precision of the study estimates. In evaluating estimates from observational studies, it is generally helpful to consider the standard deviations of estimates and the width of confidence intervals. A large standard deviation relative to the estimate indicates low precision. Similarly, wide confidence intervals for estimates of association (e.g., odds ratios or relative risks) indicate low precision.

Evaluating the Validity of an Observational Study

A second general challenge of observational research is validity. Whereas precision is a lack of random error, validity refers to a lack of systematic error.[5] Observational studies are evaluated in terms of both internal and external validity. Internal validity refers to the strength of the inferences from the study. That is, did the "exposure" or "intervention" cause a difference in the outcome (high internal validity) or was a difference in the outcome caused by systematic error in the study (low internal validity). The key question in assessing internal validity is whether observed changes can be attributed to the exposure and not to other possible causes. The internal validity of a study may be compromised by not having a control group or by having a control group that is not comparable to the exposed group in measurable or unmeasurable ways.

External validity is the ability to generalize study results to a more universal population.[5] Inferences about cause–effect associations from a specific study are considered externally valid if they may be generalized from the unique and idiosyncratic settings, procedures and participants of the study, to other populations and conditions. External validity is the degree to which the conclusions in a study would hold for other persons in other places and at other times. As such, internal validity is a prerequisite for external validity. That is, the study must demonstrate that the "exposure" in the study is the cause of variation in the outcome before one can generalize that the exposure more universally causes the outcome.

One indication that a study lacks external validity is if the sample is not representative. The most common loss of external validity in observational research comes from the fact that studies often employ small samples obtained from a single geographic location or facility. Because of this, one cannot be sure that the conclusions drawn about cause-effect-relationships apply to people in other geographic locations or at other facilities. The best way for the field of palliative care to demonstrate external validity of research results is to replicate results in different populations, places, and time periods.

Conclusions

Experimental and observational research methods are complementary tools that each plays a vital role in understanding and improving palliative care. Well-designed observational and quasi-experimental studies can provide valuable new knowledge that will advance the field of palliative care. Nevertheless, the limitations of observational research require that investigators and palliative care practitioners be critically aware of the pitfalls of these types of designs and ensure that they are appropriately recognized and addressed.

Questions a Critical Reader Should Ask

1) Is the study design appropriate for the research question?
2) Is the sample size adequate and evenly distributed across groups and covariates?
3) Are the groups comparable aside from exposure status?
4) Do estimates appear precise?
5) Is causality appropriately attributed?
6) Is there potential bias due to differential loss to follow up?
7) Is there potential recall or interviewer bias?
8) Have alternative explanations been considered and ruled out?
9) Are results placed in the context of existing literature and inconsistencies explained?
10) Is there external validity?

Acknowledgments

Dr. Carlson is a Brookdale National Fellow, and Olive Branch Scholar of the National Palliative Care Research Center, and the recipient of a Pathway to Independence Award (K99NR10495) from the National Institute for Nursing Research.

Dr. Morrison is the recipient of a Mid-Career Investigator Award in Patient Oriented Research from the National Institute on Aging (K24AG022345).

References

1. Carlson MD, Morrison RS: User's guide to research in palliative care: Why is a new series needed? J Palliat Med 2008;11:1258–1261.
2. Black N: Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312:1215–1218.
3. Mamdani M, Sykora K, Li P, Normand SL, Streiner DL, Austin PC, Rochon PA, Anderson GM: Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. BMJ 2005;330:960–962.
4. Rochon PA, Gurwitz JH, Sykora K, Mamdani M, Streiner DL, Garfinkel S, Normand SL, Anderson GM: Reader's guide to critical appraisal of cohort studies: 1. Role and design. BMJ 2005;330:895–897.
5. Rothman KJ, Greenland S: *Modern Epidemiology, Second Edition*. Philadelphia, PA: Lippincott-Raven, 1998.
6. Bausewein C, Higginson IJ: Appropriate methods to assess the effectiveness and efficacy of treatments or interventions to control cancer pain. J Palliat Med 2004;7:423–430.
7. Carlson MD, Morrison RS, Holford TR, Bradley EH: Hospice care: What services do patients and their families receive? Health Serv Res 2007;42:1672–1690.
8. Carlson MD, Gallo WT, Bradley EH: Ownership status and patterns of care in hospice: Results from the National Home and Hospice Care Survey. Med Care 2004;42:432–438.
9. Goldstein NE, Concato J, Fried TR, Kasl SV, Johnson-Hurzeler R, Bradley EH: Factors associated with caregiver burden among caregivers of terminally ill patients with cancer. J Palliat Care 2004;20:38–43.
10. Bradley EH, Fried TR, Kasl SV, Cicchetti DV, Johnson-Hurzeler R, Horwitz SM: Referral of terminally ill patients for hospice: Frequency and correlates. J Palliat Care 2000; 16:20–26.
11. Writing Group for the Women's Health Initiative Investigators: Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women's Health Initiative Randomized Control Trial. JAMA 2002;288:321–333.
12. Gelfman LP, Meier DE, Morrison RS: Does palliative care improve quality? A survey of bereaved family members. J Pain Symptom Manage 2008;36:22–28.
13. Morrison RS, Magaziner J, Gilbert M, Koval KJ, McLaughlin MA, Orosz G, Strauss E, Siu ALL: Relationship between pain and opioid analgesics on the development of delirium following hip fracture. J Gerontol A Biol Sci Med Sci 2003;58:76–81.
14. Krauss MJ, Evanoff B, Hitcho E, Ngugi KE, Dunagan WC, Fischer I, Birge S, Johnson S, Costantinou E, Fraser VJ: A case-control study of patient, medication, and care-related risk factors for inpatient falls. J Gen Intern Med 2005;20: 116–122.

Address reprint requests to:
Melissa D.A. Carlson, Ph.D.
Box 1070
Mount Sinai School of Medicine
New York, NY 10029

E-mail: melissa.carlson@mssm.edu